

REMARKS

Claims 1-22 are pending in the present application. Claims 1-3, 7, 9, 10, 11, 13, 15, 18 and 20 are amended. Reconsideration of the claim rejections are respectfully requested in view of the following remarks.

Claim Rejections - § 103

Claims 1-22 stand rejected under 35 U.S.C 103(a) as being unpatentable over U.S. Patent No. 5,995,930 to Hab-Umbach in view of U.S. Patent No. 6,078,885 to Beutnagel, as set forth in pages 2-3 of the Office Action.

The claims have been amended, inter alia, to replace instances of N text sequences with N textual transcriptions to clarify that the synthetic waveforms are generated from text that was automatically transcribed from an input waveform using a speech recognition system as opposed to plain text, which may have been manually entered without any relationship to its audio counterpart. N has been further described as being greater than 1 because at least 2 textual transcription guesses of an input waveform are generated to enable a comparison of their corresponding synthetic waveforms.

It is respectfully submitted that Hab-Umbach and Beutnagel, alone or in combination, do not disclose or suggest *generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent N-best textual transcription hypotheses of the original waveform*, as recited in amended claim 1.

While Hab-Umbach discloses test signals and references signals, neither of these signals are waveforms generated from textual transcriptions output by a speech recognition system. With respect to the test signals, Hab-Umbach merely teaches (in col., 5, lines 20-30) that the test signals are generated from a continuous speech signal obtained from a

microphone. However, Hab-Umbach fails to teach the continuous speech signal being a transcription or the continuous speech signal being obtained from a speech recognition system. With respect to the reference signals, Hab-Umbach merely teaches (in col. 5, lines 28-29 and col. 8, lines 29-30) that the reference signals are pre-defined in memories 24 or 116, and does **not** teach how they are generated. Indeed, the references signals are only discussed with respect to what they form and **not** how they are generated. For example, Hab-Umbach teaches (in the abstract) that a series of references signals form (generate) one of a plurality of vocabulary words arranged as a vocabulary tree. Hab-Umbach further teaches (in col. 57-62, lines and FIG. 1) that branches of the tree correspond units of speech **sound** (phonemes). Hab-Umbach arguably teaches generation of units of speech **sounds** from a series of the reference signals, but is limited to the content of sound; Hab-Umbach does not teach or suggest methods for processing text. Claim 1 essentially recites generation of a waveform for each of N textual transcriptions. Even assuming *arguendo* that a series of connected branches of phonemes in the tree are interpreted as a waveform, the phonemes are generated from the reference signals and there is **no** teaching in Hab-Umbach of the reference signals being textual transcriptions.

Further, the Examiner concedes (in p. 2 of the Office Action) that Hab-Umbach does not specifically teach generation of a synthetic waveform for each of N-best text sequences. Accordingly, it follows that Hab-Umbach also does not teach generation of a synthetic waveform for each of N textual transcriptions.

Moreover, the deficiencies of Hab-Umbach in regard to *generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition*

system and represent N-best textual transcription hypotheses of the original waveform is not cured by Beutnagel. Beutnagel merely teaches (in col. 4, lines 21-25 and FIG. 2) generation of N candidate pronunciations for a single given written word. However, Beutnagel does not disclose the text being generated by a speech recognition system as a textual transcription of an original waveform.

Further, even assuming *arguendo* that the single given word is interpreted as a textual transcription of an original waveform and a candidate pronunciation is interpreted as a synthetic waveform, generation of N synthetic waveforms for 1 textual transcription does not disclose generation of a synthetic waveform for each of N (e.g., 2) textual transcriptions. The following example further clarifies this distinction. In claim 1, N textual transcriptions (e.g., 2) of an input waveform from a speech recognition system are present and a synthetic waveform for each the N textual transcriptions (e.g., 2) is generated. For example, assume a speech recognition system determined that an input waveform is more likely to resolve to textual transcriptions of “Peabody” and “Prebody” and then 2 synthetic waveforms would be created (e.g., synPeabody and synPrebody). However, Beutnagel teaches (in FIG. 2) a human operator only selecting a single given word at a time. For example, in Beutnagel a user could only choose either “Peabody” or “Prebody” (e.g., assume “Peabody”) and only a pronunciation for the chosen word would be generated by the system. Thus, even if that pronunciation were interpreted as a synthetic waveform (e.g., synPeabody), the system of Beutnagel does not also generate a synthetic waveform for “Prebody” (e.g., “Peabody”).

Moreover, the Examiner asserts, without support, that “it was well known in the art to implement generating synthetic speech of N-best text sequences for the purpose of

reducing recognition errors due to decoding errors of acoustically similar words”.

However, it is not believed that use of synthetic speech to reduce recognition errors is well-known since synthetic speech can cause errors in speech recognition. For example, the text of “pinelawn” can produce synthetic speech that sounds like “pinny lawn” when “pinelawn” is broken down into phenomes “pin”, “e”, and “lawn”. Further, the Examiner does not include support for the above assertion. As stated in MPEP 2144.03, “[i]f the examiner is relying on personal knowledge to support the finding of what is known in the art, the examiner must provide an affidavit or declaration setting forth specific factual statements and explanation to support the finding. The Examiner has provided no such affidavit or declaration.

For at least the foregoing reasons, claim 1 is believed to be patentable over Hab-Umbach and Beutnagel. Amended independent claims 9 and 15 are believed to be patentable over Hab-Umbach and Beutnagel for at least similar reasons. The claims that depend from claims 1, 9, and 15 are believed to be patentable over Hab-Umbach and Beutnagel at least by virtue of their dependence to corresponding base claims. Withdrawal of the rejections under 103(a) is respectfully requested.

In view of the foregoing remarks, it is respectfully submitted that all the claims now pending in the application are in condition for allowance. Early and favorable reconsideration is respectfully requested.

Respectfully submitted,

By: Robert Newman
Robert Newman
Reg. No. 60,718
Attorney for Applicants

F. CHAU & ASSOCIATES, LLC
130 Woodbury Road
Woodbury, NY 11797
Telephone: (516) 692-8888
Facsimile: (516) 692-8889